Feature Selection Considering Multiple Correlations Based on Soft Fuzzy Dominance Rough Sets for Monotonic Classification

Abstract—This article talks about a common task in decision-making called monotonic classification. It explains that there is a mathematical tool called dominance-based rough set theory that can help with this task, but it is sensitive to noisy information and can lead to errors. The article proposes a new model called robust fuzzy dominance rough set and a feature selection method that takes into account multiple correlations between features. The article also reports on experiments done on 12 datasets and shows that the proposed model and feature selection algorithm work well for classification tasks.

I. INTRODUCTION

This passage discusses Monotonic classification tasks (MCTs), which are common in decision analysis for tasks like credit rating assessment and financial risk prediction. MCTs have a monotonic constraint between features and decisions, meaning that if a sample's feature values follow a certain order, then its decision values must also follow that same order. While there has been significant research on MCTs in recent years, there are few studies on feature selection methods for these tasks.

This passage discusses the use of fuzzy rough sets theory as a powerful tool for computing the fuzziness and uncertainty of classification tasks. To combat noise information, several robust fuzzy rough set (FRS) models have been proposed, including β-precision fuzzy rough sets, variable precision fuzzy rough sets, vaguely quantified rough sets, fuzzy variable precision rough sets, and soft distance-based fuzzy rough sets. However, these robust models are not suitable for MCTs because they do not take into account the monotonic constraints between samples and decisions.

The dominance-based rough set approach (DRSA) has been successfully applied to monotonic classification tasks (MCTs) as a mathematical framework to handle uncertainty. In the past two decades, various extended DRSA models have been studied, including the fuzzy dominance rough set (FDRS) model proposed by Hu et al. The FDRS model comprehensively considers the uncertainty of MCTs from two terms of fuzziness and roughness, making it useful for accurately capturing uncertainty from numerical and fuzzy MCTs. Further research on this model is necessary due to its wide application range.

The article discusses the problem of noise information in the process of data collection and its impact on classification tasks. The Fuzzy Dominance Rough Set (FDRS) model is presented as a framework that handles uncertainty in monotonic classification tasks. However, the FDRS model is not robust to mislabeled samples, which limits its real-world application. Therefore, the article aims to construct a robust FDRS model to handle noise information in classification tasks.

The article discusses the importance of feature selection in classification tasks and how it can improve performance and avoid overfitting. There are relatively few feature selection methods dedicated to MCTs, but several have been proposed in the literature. Most of these methods use different dominance rough sets as the theoretical basis and employ dependence functions combined with forward greedy search or heuristic algorithms to obtain feature subsets. However, these methods have two drawbacks: they use non-robust dominance-based rough set models and only consider the relevance of features to decision, ignoring multiple correlations between features, including redundancy, complementarity, and interaction.

The article discusses the importance of feature selection in classification tasks and mentions various feature selection methods proposed by researchers for MCTs. However, most of these methods only consider the relevance of features to the decision and do not take into account the multiple correlations between features, including redundancy, interaction, and complementarity. The article then goes on to describe different approaches to address these limitations, such as feature selection algorithms based on relevance, redundancy, and complementarity proposed by Li et al., feature selection methods considering feature interaction and complementarity in the fuzzy approximate space developed by Wan et al., and feature selection methods based on feature interactions proposed by Zhou et al. and Tang et al.

The existing feature selection methods that consider feature correlations have two major shortcomings. Firstly, most of them only consider the impact of feature interaction and complementarity for data classification but do not comprehensively consider the multiple correlations including relevance, redundancy, interaction, and complementarity. Secondly, they do not consider the monotonic classification of data, making them unsuitable for MCTs. However, in practical applications, MCTs often exhibit multiple correlations between features, which makes it difficult for independent criteria to provide useful information for decision-making. Therefore, a novel feature selection method that comprehensively considers the relevance, redundancy, interaction, and complementarity between features is necessary for MCTs.

Through the above analysis and discussion of the shortcomings of existing feature selection methods for MCTs, this study proposes a four-in-one feature selection approach based on a robust fuzzy dominance rough set for MCTs. The main contributions of this study can be summarized as follows.

1) The reason why FDRS is sensitive to noise and the influence of noise samples for approximate calculations are discussed in detail. Then, we propose an SFDRS with robustness, and introduce a feature selection algorithm with SFDRS-based dependency function (FS-SFDRS).

2) In the framework of the SFDRS, a four-in-one feature evaluation index used to evaluate the importance of features is constructed, which comprehensively considers four aspects of correlations including the relevance, redundancy, interaction, and complementarity.

3) On the basis of the above researches, a novel feature selection algorithm (i.e., SFOFS) is designed, and its time complexity is analyzed. In addition, a learning algorithm to find the optimal parameter α in the algorithms SFOFS and FS-SFDRS is designed. 4) Experimental results show that the SFDRS has good robustness, the SFOFS algorithm has better classification performance, and the proposed algorithm is sensitive to the parameter α.